

Analisis Komparasi Algoritma Naïve Bayes dan C4.5 untuk Prediksi Registrasi Mahasiswa di Universitas Dian Nuswantoro

Budy Santoso¹

^{1,3} Jurusan Teknik Informatika, FASILKOM UDINUS
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA

¹111201105986@mhs.dinus.ac.id

Abstract— The problem of poverty in Indonesia is very urgent to be addressed. Pacinan Village is one of villages whose inhabitants are classified as below-average income. Home quality especially low-income, still apprehensive such as the house walls from braided bamboo, the floor is not waterproof, there is no ventilation. Determining who is really to get a house, side of the village still make decisions like. So the assistance fallen wide of the mark. In this research, used the system applications to support the decision. It used to Weighted Product (WP) method. The goal of the reseach in order the population get a house. The implementation on this reseach using Visual Basic 6.0 programme language. The results is the highest weight value show that the population reserve to get this house.

Keywords— Decision Support System, Weighted Product (WP) Method, Visual Basic 6.0

I. PENDAHULUAN

Perguruan tinggi merupakan salah satu penyelenggara pendidikan akademi mahasiswa, ada lima lembaga perguruan tinggi yakni universitas, institut, sekolah tinggi, akademi serta politeknik [1]. Menurut data resmi Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan Republik Indonesia, saat ini sudah ada 100 Perguruan Tinggi Negeri di Indonesia, berbentuk universitas, institut, sekolah tinggi, akademi serta politeknik dan terdapat 3.078 Perguruan Tinggi Swasta yang tersebar dari Aceh Sampai Papua. Dengan data perguruan tinggi swasta yang sedemikian banyak, terjadi persaingan untuk menarik minat mahasiswa baru. Apabila sebuah perguruan tinggi tanpa mahasiswa baru setiap tahunnya, berarti perguruan tinggi tersebut terancam tutup. Setiap perguruan tinggi ingin mendapatkan mahasiswa yang memiliki kualitas yang baik serta dengan dengan kuantitas maksimal sesuai dengan kuota yang di tetapkan dari perguruan tinggi tersebut. Untuk perguruan tinggi negeri ada Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) yang mempermudah untuk mendapat mahasiswa baru. Berbeda dengan perguruan tinggi swasta yang harus berusaha lebih untuk mendapatkan mahasiswa baru. Maka perguruan tinggi harus memiliki strategi agar bisa bersaing dalam menarik minat mahasiswa baru. Banyak macam strategi yang digunakan oleh perguruan tinggi swasta, antara lain dengan membuka pendaftaran mahasiswa baru lebih awal.

Secara garis besar ada dua tahap menjadi mahasiswa baru di sebuah perguruan tinggi yaitu tahap pendaftaran dan tahap registrasi. Perlu diketahui bahwa tidak semua calon mahasiswa baru yang melakukan pendaftaran akan melakukan registrasi khususnya di perguruan tinggi swasta. Setiap perguruan tinggi memiliki gudang data yang yang menampung berbagai macam data,

misalnya universitas dian nuswantoro yang memiliki data pendaftaran mahasiswa baru pada tahun ajaran 2014/2015 sebanyak 3711 pendaftar dan 2367 atau 63,78% mahasiswa melakukan registrasi. Dengan data mining, data tersebut dapat dimanfaatkan untuk memprediksi mahasiswa yang akan melakukan registrasi.

Knowledge discovery in database (KDD) atau yang sering disebut dengan data mining merupakan sebuah kegiatan antara lain pengumpulan data dan pemakaian data history untuk menemukan suatu pola atau keteraturan. Sederhananya data mining ini dapat menggali informasi dari gudang data dengan menggunakan metode tertentu untuk mendapat informasi atau pengetahuan baru. Oleh karena itu data mining ini bisa digunakan untuk memperbaiki pengambilan keputusan di masa depan. Ada beberapa model datamining salah satunya klasifikasi. Klasifikasi ini menggunakan data dengan target (class/label) yang berupa nilai kategorikal/nominal. Dalam perkembangannya klasifikasi ada beberapa algoritma antara lain naïve bayes dan c4.5 yang masing masing memiliki akurasi yang berbeda. Naïve bayes merupakan suatu classifier probabilitas yang sederhana didasarkan pada pengaplikasian teorema bayes[3]. Sedangkan c4.5 dikembangkan oleh J.R Quinlan yang merupakan perkembangan dari decision tree yang [3]. Decision tree mengubah bentuk data menjadi pohon keputusan kemudian menjadi sebuah rule. Pada penelitian untuk prediksi mahasiswa non aktif decision tree memiliki akurasi lebih tinggi daripada naïve bayes[1]. Namun pada penelitian lain, naïve bayes memiliki akurasi lebih tinggi dari descision tree dalam klasifikasi berkaitan dengan web page [2]. Penulis berkesimpulan bahwa masing masing algoritma memiliki akurasi yang berbeda tergantung pada data yang di klasifikasi.

Maka penulis akan membandingkan algoritma naïve bayes dan c4.5 yang merupakan perkembangan dari algoritma decision tree untuk memprediksi data registrasi mahasiswa baru di universitas dian nuswantoro. Data akan diklasifikasi menggunakan algoritma naïve bayes dan c4.5, Sehingga universitas dian nuswantoro bisa mempersiapkan strategi yang tepat untuk memaksimalkan mahasiswa baru yang melakukan registrasi.

II. STUDI PUSTAKA

2.1. Penelitian Terkait

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai algoritma yang akan digunakan penulis.

1. Khafiizh Hastuti "Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif".
2. Devi Sugianti "Algoritma Bayesian Classification Untuk Memprediksi Heregistrasi Mahasiswa Baru Di Stmik Widya Pratama".
3. Alifi adia pranata "Analisis Perbandingan Lima Metode Klasifikasi pada Dataset Sensus Penduduk".

2.2. Tinjauan Pustaka

A. Data Mining

Data mining adalah melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data [5]. Dalam buku yang di tulis oleh bramer di jelaskan bahwa proses dari data mining adalah data masuk dari berbagai sumber, hal ini terintegrasi dan ditempatkan di tempat penyimpanan data atau bisa di sebut database. kemudian diambil dan dilakukan analisa. Setelah itu diteruskan ke algoritma data mining yang menghasilkan output dalam bentuk aturan atau beberapa jenis lain dari 'pola'. Ini kemudian diinterpretasikan untuk memberikan pengetahuan Penemuan baru dan pengetahuan berpotensi akan berguna [7]

B. Algoritma Naïve Bayes

Metode Klasifikasi adalah Salah satu dari beberapa tugas yang dapat dilakukan Data Mining, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yaang sudah didefinisikan sebelumnya. Salah satu metoda dalam klasifikasi data adalah Naïve Bayes. Naïve Bayes merupakan salah satu metoda machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Teorema Bayes adalah suatu pendekatan pada ketidaktentuan yang diukur dengan probabilitas [9]. Teorema Bayes memiliki rumusan umum sebagai berikut:

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)}$$

Dimana :

X = data dengan class yang belum diketahui

H = hipotesis data X, merupakan suatu class spesifik

$P(H|X)$ = probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)

$P(H)$ = probabilitas hipotesis H (prior probability)

$P(X|H)$ = probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ =probabilitas dari X

C. Algoritma C4.5

Di akhir tahun 1970 hingga di awal tahun1980-an, Quinlan[5] seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (Iterative Dichotomiser), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis supervised learning. Salah satu metode klasifikasi menarik yang melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar samapai berakhir di node daun. Dimulai dari node root, yang oleh konvensi ditempatkan dibagian atas dari diagram pohon keputusan, atribut diuji pada node keputusan, dengan setiap hasil yang mungkin menghasilkan cabang. Berikut adalah rumus pada algoritma C4.5.

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p_i)$$

$$Gain(A) = Info(D) - Info_A(D)$$

$$split_point = \frac{a_i + a_{i+1}}{2}$$

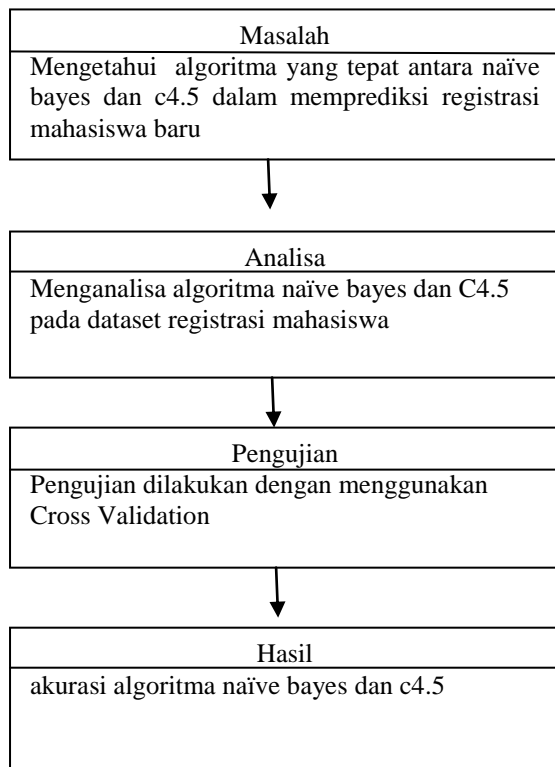
Dimana :

D = Dataset

m = banyaknya partisi D

pi = probabilitas

2.3. Kerangka Pemikiran



III METODE PENELITIAN

3.1 Instrumen Penelitian

Berdasarkan permasalahan yang telah diuraikan pada bab sebelumnya, maka bahan dan peralatan yang diperlukan untuk penelitian ini meliputi :

A. Bahan

Dalam penelitian ini bahan yang dibutuhkan adalah data pendaftaran mahasiswa baru Universitas Dian Nuswantoro tahun 2013/2014 dan 2014/2015.

B. Peralatan

Peralatan dalam penelitian ini meliputi kebutuhan perangkat lunak dan kebutuhan perangkat keras. Prototipe akan menggunakan bahasa pemrograman php dan menggunakan database mysql sehingga membutuhkan perangkat lunak xampp dan sqlyog. Dibawah ini merupakan sistem yang akan digunakan dalam membuat prototype:

Kebutuhan Perangkat Lunak

- Sistem Operasi Windows 7 Ultimate
- Sistem Type 64-bit Operating System

Kebutuhan perangkat keras

- Processor Intel(R) Core(TM) i3-2310M CPU @2.10GHz 2.1 GHz
- Ram 2GB
- Harddisk 500GB
- Satu buah mouse

3.2 Pengumpulan Data

Data yang digunakan pada penelitian ini berasal dari data pendaftaran mahasiswa baru Universitas Dian Nuswantoro tahun 2013/2014 dan 2014/2015, data tersebut di peroleh dari universitas dian nuswantoro dengan 5 variabel.

3.3 Teknik Analisis Data

Data mahasiswa baru Universitas Dian Nuswantoro tahun 2013/2014 dan 2014/2015 yang diperoleh dari pengumpulan data sebanyak 8771 record, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (preparation data). Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut:

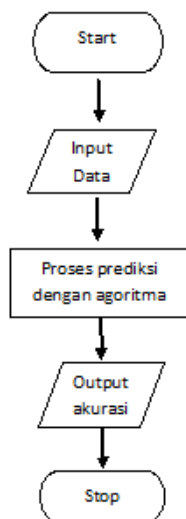
- 1.Data cleaning.
- 2.Data integration and transformation
- 3.Data reduction

Tabel 3.1 atribut

	Nama atribut	Penjelasan
1	Program studi	Program studi yang dipilih oleh mahasiswa baru
2	Fakultas	Fakultas yang dipilih oleh mahasiswa baru
3	Kota asal	Kota asal mahasiswa baru
4	Gelombang	Gelombang pendaftaran mahasiswa baru
5	Status registrasi	Status apakah mahasiswa baru mrlakukan registrasi atau tidak

3.4 Eksperimen

Pada tahap akhir pada penelitian ini akan dilakukan pengujian pada eksperimen yang telah dilakukan. Pada tahap eksperimen ini penulis menerapkan model yang telah diusulkan mejadi sebuah prototype yang di bangun menggunakan php dan mysql.



IV HASIL PENELITIAN DAN PEMBAHASAN

Didalam bab ini akan dibahas mengenai data yang akan digunakan dalam penelitian, data tersebut akan di hitung menggunakan algoritma naïve bayes dan C4.5 kemudian diuji menggunakan cross validation.

4.1 Data yang digunakan

Pada penelitian ini, data yang digunakan adlah data registrasi mahasiswa universitas dian nuswantoro tahun 2013/214 dan 2014/2015 dengan jumlah data 3624 record.

Tabel 4. 1 Tabel Data Registrasi Mahasiswa

progdi	fakultas	kota_asal	gelombang	status
Akuntansi	Ekonomi & Bisni	luar kota	IIB	Registrasi
Akuntansi	Ekonomi & Bisni	luar kota	PMDKIII	Registrasi
Akuntansi	Ekonomi & Bisni	luar kota	IIA	Registrasi
Akuntansi	Ekonomi & Bisni	dalam kota	PMDKII	Tidak Registras
Desain Komunikasi Vi	Ilmu Komputer	luar kota	PMDKII	Registrasi
Desain Komunikasi Vi	Ilmu Komputer	dalam kota	IIB	Registrasi
Desain Komunikasi Vi	Ilmu Komputer	dalam kota	PMDKII	Tidak Registras
Broadcasting	Ilmu Komputer	luar kota	IA	Registrasi
Broadcasting	Ilmu Komputer	luar kota	IIB	Registrasi
Broadcasting	Ilmu Komputer	dalam kota	IC	Registrasi

4.2 Perhitungan Algoritma Naïve Bayes

Dibawah ini adalah contoh perhitungan manual penerapan algoritma naïve bayes untuk memprediksi registrasi mahasiswa menggunakan data training dan data esting sebagai berikut:

Tabel 4.2 Perhitungan Algoritma Naïve Bayes

progdi	fakultas	Kota Asal	Gelombang	Status
akuntansi	Ekonomi & Bisnis	Dalam Kota	IC	?

Dalam perhitungan algoritma naïve bayes ada beberapa langkah yang harus dilakukan, berikut adalah langkah langkahnya

- Pertama harus menghitung jumlah class/ label
 $P(\text{status} = \text{registrasi}) = 2365/3624 = 0,65$
 $P(\text{status} = \text{tidak registrasi}) = 1259/3624 = 0,35$
- Yang kedua adalah menghitung jumlah kasus yang sama dengan class/label yang sama
 $P(\text{progdi} = \text{akuntansi} \mid \text{status} = \text{registrasi}) = 226/2365 = 0,095$
 $P(\text{progdi} = \text{akuntansi} \mid \text{status} = \text{tidak registrasi}) = 132/1259 = 0,104$
 $P(\text{fakultas} = \text{Ekonomi \& Bisnis} \mid \text{status} = \text{registrasi}) = 596/2365 = 0,25$
 $P(\text{fakultas} = \text{Ekonomi \& Bisnis} \mid \text{status} = \text{tidak registrasi}) = 334/1259 = 0,26$
 $P(\text{kota asal} = \text{dalam kota} \mid \text{status} = \text{registrasi}) = 106/2365 = 0,044$
 $P(\text{kota asal} = \text{dalam kota} \mid \text{status} = \text{tidak registrasi}) = 55/1259 = 0,043$
 $P(\text{gelombang} = \text{IC} \mid \text{status} = \text{registrasi}) = 358/2365 = 0,151$
 $P(\text{gelombang} = \text{IC} \mid \text{status} = \text{tidak registrasi}) = 235/1259 = 0,186$

- Yang terakhir lakukan perkalian masing masing hasil perhitungan kasus yang sama dengan class/label yang sama dengan memperhatikan class/label. Setelah itu bandingkan hasilnya.
 $P(X \mid \text{status} = \text{registrasi}) P(\text{status} = \text{registrasi}) = 0.095 * 0.025 * 0.044 * 0.151 * 0.65 = 0,000106$
 $P(X \mid \text{status} = \text{tidak registrasi}) P(\text{status} = \text{tidak registrasi}) = 0,104 * 0,26 * 0,043 * 0,186 * 0,35 = 0,0000787$

Karena probabilitas registrasi lebih beasr, maka kesimpulannya adalah pendaftar dengan progdi Akuntansi, fakultas Ekonomi & Bisnis, kota asal Dalam Kota, gelombang IC akan melakukan registrasi

4.3 Perhitungan Algoritma C4.5

Perhitungan menggunakan algoritma C4.5 berbeda dengan algoritma naïve bayes. Pada Algoritma C4.5 harus menentukan pohon keputusan yang kemudian akan menjadi rule untuk memprediksi registrasi mahasiswa.

Tabel.4.3 literasi 1

atribut	nilai atribut	jumlah	regist	tidak regis	entropy	gain
total		3624	2365	1259	0.931730198	0.00819435
progdi	akuntansi	358	226	132	0.949680388	
	broadcast	69	45	24	0.932111568	
	dkv	279	183	96	0.928675181	
	ilmu kom	125	100	25	0.721928095	
	kesmas	245	160	85	0.931304369	
	manajemen	572	370	202	0.936847106	
	manajemen informatika	20	14	6	0.881290899	
	rekam medis	167	135	32	0.704841385	
	sastra inggris	91	60	31	0.925448252	
	sastra jepang	60	39	21	0.934068055	
	sistem informasi	268	179	89	0.917047028	
	teknik elektro	72	50	22	0.88797632	
	teknik industri	148	89	59	0.970154614	
	teknik informatika	1150	715	435	0.956804463	
fakultas	ekonomi dan bisnis	930	596	334	0.941966829	0.00175545
	ilmu budaya	151	99	52	0.928940032	
	ilmu komputer	1911	1236	675	0.936909397	
	kesehatan	412	295	117	0.860818101	
	teknik	220	139	81	0.949265807	
gelombang IA		133	89	44	0.915769153	0.07023024
	IB	291	230	61	0.740759572	
	IC	593	358	235	0.968739002	
	IIA	345	270	75	0.755375413	
	IIB	1084	858	226	0.738580888	
	KHUSUS	38	25	13	0.926819064	
	PMDK1	204	64	140	0.897427193	
	PMDK2	385	214	171	0.990982908	
	PMDK3	551	257	294	0.996744845	
KOTA ASAL	dalam kota semarang	1341	916	425	0.901009064	0.59832701
	luar kota semarang	2284	1449	834	0.947218735	

Berdasarkan tabel diatas, berikut adalah contoh perhitungan mencari entropy dan gain pada field progdi menggunakan algoritma C4.5

$$\text{Entropy (total)} = (- (2365/3624) * (\log_2 (2365/3624))) + (- (1259/3624) * (\log_2 (1259/3624))) = 0,9317$$

$$\text{Entropy (progdi (akuntansi))} = (- (226/358) * (\log_2 (226/358))) + ((132/358) * (\log_2 (132/358))) = 0,9496$$

$$\text{Entropy (progdi (broadcast))} = (- (45/69) * (\log_2 (45/69))) + ((24/69) * (\log_2 (24/69))) = 0,9321$$

$$\text{Entropy (progdi (dkv))} = (- (183/279) * (\log_2 (183/279))) + ((96/279) * (\log_2 (96/279))) = 0,9286$$

$$\text{Entropy (progdi (ilmu kom))} = (- (100/125) * (\log_2 (100/125))) + ((25/125) * (\log_2 (25/125))) = 0,7219$$

$$\text{Entropy (progdi (kesmas))} = (- (160/245) * (\log_2 (160/245))) + ((85/245) * (\log_2 (85/245))) = 0,9313$$

$$\text{Entropy (progdi (manajemen))} = (- (370/572) * (\log_2 (370/572))) + ((202/572) * (\log_2 (202/572))) = 0,9313$$

$$\text{Entropy (progdi (manajemen informatika))} = (- (14/20) * (\log_2 (14/20))) + ((6/20) * (\log_2 (6/20))) = 0,8812$$

$$\text{Entropy (progdi (Rekam medis))} = (- (135/167) * (\log_2 (135/167))) + ((32/167) * (\log_2 (32/167))) = 0,7048$$

$$\text{Entropy (progdi (sastra inggris))} = (- (60/91) * (\log_2 (60/91))) + ((31/91) * (\log_2 (31/91))) = 0,9254$$

$$\text{Entropy (progdi (sastra jepang))} = (- (39/60) * (\log_2 (39/60))) + ((21/60) * (\log_2 (21/60))) = 0,9340$$

$$\text{Entropy (progdi (sistem informasi))} = (- (179/268) * (\log_2 (179/268))) + ((89/268) * (\log_2 (89/268))) = 0,9170$$

$$\text{Entropy (progdi (teknik elektro))} = (- (50/72) * (\log_2 (50/72))) + ((22/72) * (\log_2 (22/72))) = 0,8879$$

$$\text{Entropy (progdi (teknik industri))} = (- (89/148) * (\log_2 (89/148))) + ((59/148) * (\log_2 (59/148))) = 0,9701$$

$$\text{Entropy (progdi (teknik informatika))} = (- (715/1150) * (\log_2 (715/1150))) + ((435/1150) * (\log_2 (435/1150))) = 0,9568$$

$$\begin{aligned} \text{Gain (Progdi)} &= 0,9317 - ((358/3624) * 0,9496) + \\ &((69/3624) * 0,9321) + ((279/3624) * 0,9286) + ((125/3624) * \\ &0,7219) + ((245/3624) * 0,9313) + ((572/3624) * 0,9368) + \\ &((20/3624) * 0,8812) + ((167/3624) * 0,7048) + ((91/3624) * \\ &0,9254) + ((60/3624) * 0,9340) + ((268/3624) * 0,9170) + \\ &((72/3624) * 0,8879) + ((148/3624) * 0,9701) + \\ &((1150/3624) * 0,9568) = 0,00819 \end{aligned}$$

4.4 Pengujian Cross Validation

Pada pengujian akurasi Penulis menggunakan 37 data testing, 3624 data training dan akan melakukan ujicoba sebanyak 10 kali. Data tersebut akan diuji untuk mengetahui akurasi yang dihasilkan oleh algoritma naïve bayes dan C4.5 dengan cross validation, kemudian akan di analisa, sehingga bisa diketahui jumlah benar dan salah saat melakukan prediksi pada data testing tersebut. Setelah mengetahui jumlah perbedaannya dalam melakukan proses prediksi. Dari 10 kali percobaan dengan 37 data testing yang berbeda dihasilkan 10 akurasi yang berbeda pula. Dari 10 akurasi diatas masing masing algoritma memiliki rata rata yaitu algoritma naïve bayes memiliki rata rata 82,16 % dan algoritma C4.5 memiliki rata rata 80,80 %.

4.5 Prototype

Gambar 4. 1 Implementasi

Kesimpulan :registrasi (6.8502555928453E-5)			
Naive Bayes	Registrasi	Tidak Registrasi	
progdi Akuntansi	226/2365=0.095560253699789	132/1259=0.10484511517077	
fakultas Ekonomi & Bisnis	596/2365=0.25200845665962	334/1259=0.26528991262907	
kota asal Dalam Kota	106/2365=0.044820295983087	55/1259=0.043685464654488	
gelombang IB	230/2365=0.097251585623679	61/1259=0.048451151707705	
probabilitas	6.8502555928453E-5	2.0452553774428E-5	

Gambar 4.2 Implementasi naïve bayes

progdi	Manajemen
fakultas	Ekonomi & Bisnis
kota asal	Dalam Kota
gelombang	IA
hasil prediksi	registrasi

[Kembali](#)

Gambar 4. 3 Implementasi C4.5

V KESIMPULAN DAN SARAN

a. Kesimpulan

Berdasarkan penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan beberapa hal sebagai berikut :

1. Registrasi mahasiswa baru di universitas dian nuswantoro semarang dapat di prediksi dengan menggunakan beberapa variable yaitu progdi, fakultas, kota asal dan gelombang.
2. Dari hasil percobaan prediksi registrasi mahasiswa dengan metode naïve bayes dan c4.5 menggunakan 37 data testing dan 3624 data training diperoleh rata-rata akurasi sebesar 82,16 % untuk algoritma naïve bayes dan 80,8 % untuk algoritma c4.5. Sehingga antara algoritma naïve bayes dan algoritma c4.5 yang lebih tepat digunakan untuk memprediksi registrasi mahasiswa universitas dian nuswantoro adalah algoritma naïve bayes.
3. Dengan data registrasi mahasiswa yang diolah dengan algoritma naïve bayes dan c4.5, universitas dian nuswantoro bisa memperoleh informasi yang bisa digunakan untuk mempersiapkan strategi menarik mahasiswa baru.

b. Saran

Untuk meningkatkan kinerja dan menyempurnakan sistem pendukung keputusan yang telah dibuat, peneliti memberikan saran sebagai berikut :

1. Diharapkan dapat menambah variabel seperti pendapatan dan pekerjaan orang tua untuk menghasilkan prediksi yang lebih baik.
2. Diharapkan dapat membandingkan dengan algoritma klasifikasi yang lain agar mengetahui algoritma yang lebih baik dalam memprediksi registrasi mahasiswa.

UCAPAN TERIMA KASIH

Penulis mengucapkan teima kasih kepada Universitas Dian Nuswantoro, Rektor UDINUS, Dekan Fakultas Ilmu Komputer, Kaprodi Teknik Informatika-S1, Dosen pembimbing, Dosen-dosen pengampu kuliah di Fakultas Ilmu Komputer , serta teman-teman dan sahabat yang selama ini telah mendampingi penulis selama kuliah di Universitas Dian Nuswantoro.

REFERENSI

- [1] Khafiizh Hastuti, "Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif," Seminar Nasional Teknologi Informasi & Komunikasi Terapan (Semantik), 2012.
- [2] Xhemali, D., Hinde, C.J. And Stone, R.G. " Naïve Bayes Vs. Decision Trees Vs. Neural Networks In The Classification Of Training Web Pages", International Journal Of Computer Science Issues, 4 (1), 2009.
- [3] Alifi Aida Pranatha, "Analisis Perbandingan Lima Metode Klasifikasi pada Dataset Sensus Penduduk", Jurnal Sistem Informasi, 2012.
- [4] Devi Sugianti, "Algoritma Bayesian Classification Untuk Memprediksi Heregistrasi Mahasiswa Baru di Stmik Widya Pratama", Jurnal Ilmiah Ictech, 2012.
- [5] E. Prasetyo, Data Mining : Konsep dan Aplikasi Menggunakan Matlab, Yogyakarta: Andi , 2014.
- [6] ian H. Witten, Eibe Frank, Mark A Hall, Data Mining : Practial Machine Learning Tools and Techniques, Burlington : Morgan Kaufmann, 2011
- [7] Max Bramer, Data Mining : Principles of Data Mining, London : Springer, 2007
- [8] F. Gorunescu, Data Mining Concept Model Technique, Craiova, Romania: Springer, 2011.
- [9] Ferry, "Analisis Komparasialgoritma Klasifikasi Data Miningpada Akseptasi Data Fakultatifreasurensi Jiwa", tesis Sekolah Tinggi Manajemen Informatika Dan Komputer Eresha Jakarta, 2011.
- [10] Edy, "Penerapan Algoritma C4.5 Dengan Seleksi Atribut Berbasis Algoritma Genetika Dalam Diagnosa Penyakit Jantung", tesis Sekolah Tinggi Manajemen Informatika Dan Komputer Eresha Jakarta, 2012.
- [11] P. Chapman, CRISP-DM 1.0: Step-by-step Data Mining Guide, SPSS, 2000.

